International Journal of Bifurcation and Chaos, Vol. 13, No. 9 (2003) 2657–2668

SYMBOLIC ANALYSIS OF HIGH-DIMENSIONAL TIME SERIES

KARSTEN KELLER
Department of Mathematics and Computer Science, University Greifswald, Jahnstr. 15a, 17487 Greifswald

HEINZ LAUFFER
Department of Pediatric Medicine, University Greifswald, Soldtmannstr. 15, 17487 Greifswald

Received May 28, 2002; Revised July 15, 2002

In order to extract and to visualize qualitative information from a high-dimensional time series, we apply ideas from symbolic dynamics. Counting certain ordinal patterns in the given series, we obtain a series of matrices whose entries are symbol frequencies. This matrix series is explored by simple methods from nominal statistics and information theory. The method is applied to detect and visualize qualitative changes of EEG data related to epileptic activity.

Keywords: Time series analysis; EEG analysis; symbolic dynamics; entropy; contingency; correspondence analysis.

1. Introduction

Dynamical systems can be considered on different levels of precision. For studying the qualitative behavior of a system, it may be useful to choose a very coarse-grained description: The state space is divided into a small number of subregions, each coded by a symbol. In a discrete-time dynamical system (possibly obtained by discretization), sequences of successive states of the system are turned into symbol sequences.

Symbolic dynamics studies dynamical systems on the base of the symbol sequences obtained for a suitable partition of the state space. It has been used for the investigation of nonlinear and chaotic dynamical systems (for an overview see [Hao & Zheng, 1998]). The idea of symbolic dynamics has been applied also to the qualitative analysis of empirical time series (e.g. see [Ebeling & Nicolis, 1992; Schwarz *et al.*, 1993; Finney *et al.*, 1998; Daw *et al.*, 1998]).

Here, specific patterns in the time series are labeled by symbols or finite sequences of symbols. The process of symbolizing a time series can usually be described in the context of a *delay embedding* of the time series into a d-dimensional space (see [Takens, 1981]). Dependencies, periodicities, complexity changes, etc. in the time series are detected by analyzing symbol and word frequencies, where ideas coming from information theory play a central role (see [Herzel & Große, 1995; Ebeling & Nicolis, 1992]).

In this paper we propose the study of high-dimensional time series by means of symbolic dynamics. Our intention is to develop simple methods for a first exploration of high-dimensional time series, which are robust and are realizable by very fast and flexible algorithms, and can be taken as a starting-point for a deeper investigation. In particular, we want to quantify and to visualize temporal changes of the complexity of the time series and of similarities and dissimilarities between their components. Following an idea of Bandt and Pompe [2001], we consider partitions of d-dimensional spaces revealing the ordinal structure of one-dimensional time series. On this base, a

given multidimensional time series is componentwise turned into a multidimensional symbolic time series. By counting the symbols in a time interval of fixed length being shifted along the series we get a time series of frequency matrices. So the coarse-graining of the original time series allows us to explore it with the help of methods from multivariate nominal statistics and information theory.

The method is illustrated for 19-channel EEG data from children with epileptic seizures. In particular, the results obtained confirm evidence of a loss of complexity related to epileptic activity, described on the base of the correlation dimension (see [Grassberger & Procaccia, 1983]) by [Theiler, 1995; Accardo *et al.*, 1997; Lehnertz *et al.*, 2000]. Note that the use of correlation dimension for concrete data includes some difficulties both from the computational and the interpretational points of view, which besides our interest in a multivariate approach is one reason for developing the method.

2. The Concept of Ordinal Patterns

Suppose we are given an m-dimensional time series $(\mathbf{x}_t)_{t\in T} = ((x_t^i)_{i=1}^m)_{t\in T}$. For simplicity, we want to assume that T is the set of integers. (Then the adaptation of the following to concrete data is obvious.) In view of the application to EEG data, we call the components $(x_t^i)_{t\in T}$ of the time series the ith *channels*.

For given time delay τ and dimension $d > 1$, we are interested in "ordinal patterns" of order d generated by the delay embeddings

$$(i,s) \mapsto (x^i_{s-(d-1)\tau}, x^i_{s-(d-2)\tau}, \ldots, x^i_{s-\tau}, x^i_s)\,,$$

which for fixed channel i assign to each time s the d-dimensional vector of values at times $s, s-\tau, s-2\tau, \ldots, s-(d-1)\tau$. Clearly, the greater d is, the more information on the past is provided by the vectors obtained. By the *ordinal pattern* related to a *channel-time-pair* (i,s) we mean the permutation $(r_0, r_1, \ldots, r_{d-1})$ of $(0,1,\ldots,d-1)$ defined by

$$x^i_{s-r_{d-1}\tau} \le x^i_{s-r_{d-2}\tau} \le \cdots \le x^i_{s-r_1\tau} \le x^i_{s-r_0\tau}\,.$$

In order to get a unique result, we set $r_l < r_{l-1}$ in the case that $x^i_{s-r_l\tau} = x^i_{s-r_{l-1}\tau}$. Let us illustrate the concept of an ordinal pattern by an example.

Example. Let $\tau = 2$ and assume that x^i_{s-6} $(= x^i_{s-\mathbf{3}\tau}) = 3.1$, $x^i_{s-4}(= x^i_{s-\mathbf{2}\tau}) = 5.4$, x^i_{s-2} $(= x^i_{s-\mathbf{1}\tau}) = 3.1$ and $x^i_s(= x^i_{s-\mathbf{0}\tau}) = 2.5$. Then it holds $x^i_{s-\mathbf{0}\tau} < x^i_{s-\mathbf{1}\tau} = x^i_{s-\mathbf{3}\tau} < x^i_{s-\mathbf{2}\tau}$. Reading this inequality from the right to the left, one gets two candidates for the ordinal pattern related to (i,s), namely $(2,3,1,0)$ and $(2,1,3,0)$. According to the above setting for the case of equality, the ordinal pattern is $(2,3,1,0)$.

Ordinal patterns describe the up and down in the channels on scales corresponding to different delays τ. However, we are not interested in the ordinal patterns themselves but in their distribution on the channels. Thus we identify each of them with exactly one of the "symbols" $j = 1, 2, \ldots, n = d!$.

Remark. In the introduction we have mentioned that symbolic dynamics relies on a partition of the state space. Here we do not give the partition explicitly, but clearly each ordinal pattern describes a connected piece of $\mathbb{R}^d$, and the union of all pieces is the state space $\mathbb{R}^d$. It is important to note that ordinal patterns can be computed in a very fast way, being one main reason for choosing the special partition.

3. Analyzing Frequencies of Ordinal Patterns

3.1. *Ordinal pattern frequencies*

Consider a time window $[t-\delta+1, t]$ of size δ ending at given time t. For each channel i and for each j count all times $s \in [t-\delta+1, t]$ for which the channel-time pair (i,s) provides the ordinal pattern j. The relative frequencies p_{ij} obtained after dividing the results by $m\delta$ are summarized in the matrix

$$P_t = P_t(d,\tau,\delta) = (p_{ij})_{i,j=1}^{m,n}\,,$$

which reflects the distribution of ordinal patterns in the time between $t-\delta+1$ and t. (Clearly, $\sum_{i,j=1}^m p_{ij} = 1$.) This procedure turns the original time series into a matrix time series $(P_t)_{t\in T}$ being the base of a further analysis by use of quantities from nominal statistics and information theory. In the following two paragraphs we list all these quantities. (For a short overview, see the underlined text.)

3.2. *Entropies and contingency*

In the following let t be fixed and assume that the marginal relative frequencies $p_{\cdot j} = \sum_{i=1}^m p_{ij}$; $j = 1, 2, \ldots, n$ are not vanishing. Besides the

distribution of "pooled" ordinal patterns described by $\mathbf{r} = (p_{\cdot j})_{j=1}^{n}$ we are interested in the distributions of ordinal patterns relative to the single channels. They are given by the vectors $\mathbf{r}_i = (mp_{ij})_{j=1}^{n}$ for $i = 1, 2, \ldots, m$. Clearly, $\mathbf{r} = \frac{1}{m}\sum_{i=1}^{m} \mathbf{r}_i$. The complexity of the distributions $\mathbf{r}$ and $\mathbf{r}_i$; $i = 1, 2, \ldots, m$ is quantified by the *pooled permutation entropy* (of order d)

$$\underline{H_t} = -\sum_{j=1}^{n} p_{\cdot j} \ln p_{\cdot j}$$

and the ith *channel permutation entropies* (of order d)

$$\underline{H_t^i} = -\sum_{j=1}^{n} mp_{ij} \ln(mp_{ij}) .$$

Remark. The term *permutation entropy* (of order d) is due to Bandt and Pompe [2001] who have combined Shannon's entropy concept with that what we call ordinal patterns. It is important to mention that the permutation entropy seems to be very robust with respect to noise, as reported in [Bandt & Pompe, 2001]. Also note that for interval maps the quotient of the permutation entropy of order d and $d-1$ approaches the Kolmogorov–Sinai entropy (see [Bandt *et al.*, 2002]).

A natural quantity for measuring the level of inhomogenity between the distributions $\mathbf{r}_i$ is provided by the *contingency*

$$\underline{\varphi_t^2} = \frac{1}{m}\sum_{i,j=1}^{m,n} \frac{(mp_{ij} - p_{\cdot j})^2}{p_{\cdot j}} .$$

φ_t^2 is vanishing if and only if the distributions $\mathbf{r}_i = (mp_{ij})_{j=1}^{n}$ coincide, and if they are not too different in a certain sense, the following holds

$$H_t \approx \frac{1}{m}\sum_{i=1}^{m} H_t^i + \frac{\varphi_t^2}{2} . \qquad (1)$$

This fact is discussed in Sec. 5. Formula (1) being relevant for the subsequent application, can be interpreted as follows: The overall complexity of the time series at time t decomposes into the mean complexity of the channels and a rest quantifying inhomogenity between the channels.

3.3. *One-dimensional CA-plot*

In order to visualize similarities and dissimilarities between the distributions $\mathbf{r}_i$; $i = 1, 2, \ldots, m$ obtained at some time t we use Correspondence Analysis (CA). This method for exploring frequency matrices was developed primarily in France in the 70's, by Benzécri [1972] and his students (see [Greenacre, 1984; Benzécri, 1992]). Let us give a short introduction to CA. For more details we refer the reader to the Appendix.

Consider the weighted Euclidean norm $\|(s_j)_{j=1}^{n}\| = \sqrt{\sum_{j=1}^{n} \frac{s_j^2}{p_{\cdot j}}}$ on the n-dimensional space $\mathbb{R}^n$, allowing to write the contingency as the mean squared distance of the channel distributions from the "pooled" distribution:

$$\varphi_t^2 = \frac{1}{m}\sum_{i=1}^{m} \|\mathbf{r}_i - \mathbf{r}\|^2 . \qquad (2)$$

CA provides best possible approximations $\mathbf{r}_i^{(1)}$ of the $\mathbf{r}_i$ as points of a one-dimensional affine subspace of $(\mathbb{R}^n, \|\cdot\|)$ in the sense that $\frac{1}{m}\sum_{i=1}^{m} \|\mathbf{r}_i - \mathbf{r}_i^{(1)}\|^2$ is minimal. Such a subspace contains $\mathbf{r}$, and for the "optimal" $\mathbf{r}_i^{(1)}$; $i = 1, 2, \ldots, m$ the following "orthogonal" decomposition of φ_t^2 is valid (compare (2)):

$$\varphi_t^2 = \frac{1}{m}\sum_{i=1}^{m} \|\mathbf{r}_i^{(1)} - \mathbf{r}\|^2 + \frac{1}{m}\sum_{i=1}^{m} \|\mathbf{r}_i - \mathbf{r}_i^{(1)}\|^2 . \qquad (3)$$

Remark. On the simplex of (relative) frequency distributions, the distance provided by $\|\cdot\|$ is an analogue of the χ^2-distance between absolute frequency distributions.

A *one-dimensional CA-Plot* represents the approximation $\{\mathbf{r}_1^{(1)}, \mathbf{r}_2^{(1)}, \ldots, \mathbf{r}_m^{(1)}\}$ of the system $\{\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_m\}$ as a linear set of numbers. (For all t in some time interval, we will plot the set into a two-dimensional coordinate system, orthogonal to the horizontally drawn t-axis. This will provide curves for $i = 1, 2, \ldots, m$, as to be seen in Fig. 5.) The smaller $\underline{\rho_t^{(1)}} := \frac{1}{m}\sum_{i=1}^{m} \|\mathbf{r}_i - \mathbf{r}_i^{(1)}\|^2$ is, the better the one-dimensional CA-Plot reflects similarities and dissimilarities between the ordinal pattern distributions of the channels.

According to formula (3), $\rho_t^{(1)}$ can be interpreted as the *part of contingency not explained by dimension* 1. Note that CA provides similar approximations as given above for increasing dimension, explaining more and more of the contingency (see also the Appendix).

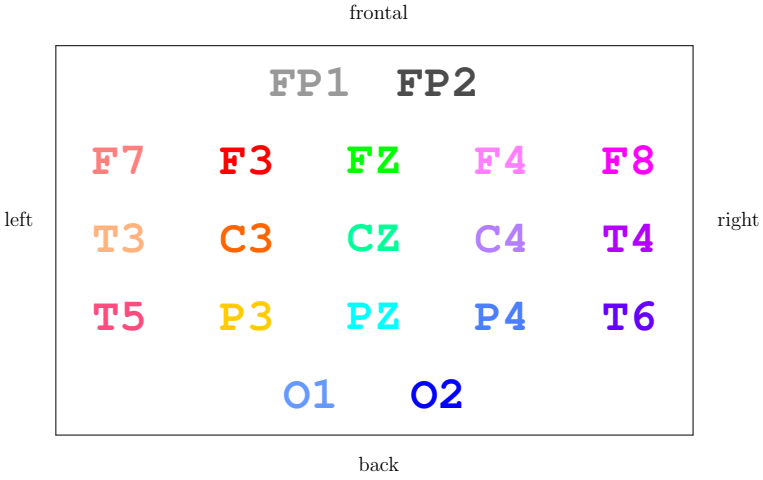

Fig. 1. 10–20-system of electrode placement.

4. Application

The method of analyzing ordinal pattern frequencies has been applied to 19-channel scalp EEG data from children with and without epileptic disorders. An EEG (electroencephalogram) records the electrical activity of the brain. In epilepsy it is used to detect and to localize abnormal brain behavior. For an introduction to EEG analysis with a special emphasis on pediatric electroencephalography we refer the reader to [Blume & Kaibara, 1999].

Our data were collected according to the international 10–20–system of electrode placement, which is illustrated in Fig. 1. The data were sampled with a rate of 256 Hz and filtered by a bandpass (0.3–70 Hz). The resulting time series has $m = 19$ components. Note that in the following figures the time is not given in data points but in seconds (one data point corresponds to 0.0039 seconds).

Good results have been obtained for the dimension $d = 4$ and for a window size of 2 sec corresponding to 512 data points. This combination, which has also been chosen in the following examples, seems to be suitable from the statistical and computational viewpoints; in particular, numerical experiments have shown that usually all $24 = 4!$ ordinal patterns are realized by the given data. Different delays displayed different details of the EEG behind the data. For example, for children without seizures the pooled permutation entropy in dependence of small delays was age-related, reflecting the development of the EEG. All measures considered were relatively stable for "normal" EEG's and parts of EEG's not related to a seizure.

On the other hand, we found strong changes of the described measures which turned out to be related to epileptic seizures, and the measures indicated differences between parts of the brain beyond the periods of epileptic discharges. Note that the complexity measured by the pooled and channel permutation entropies was strongly decreasing at the beginning of epileptic seizures, and in some examples we saw complexity changes prior to the discharges. More significant results in the latter direction have been found for intracranial EEG's by other authors using the correlation dimension (see [Lehnertz *et al.*, 2000]). Intracranial EEG's however are recorded directly from the brain providing a good signal-noise ratio, whereas the scalp EEG data contain many artefacts, e.g. caused by muscle activity. The use of correlation dimension is here particularly problematic because of its strong sensitivity with respect to noise.

Example 1. The data of an 11 year old girl with pyknoleptic absences (which are special generalized epileptic seizures seen in childhood epilepsy) are visualized in Figs. 2–5. Figure 2 shows the pooled permutation entropy over a time of 400 s for $\tau = 1, 3, 4, 5, 7$. (Note that theoretically the pooled and the channel permutation entropies cannot exceed $\ln 24 \approx 3.178$.) Four intervals of low entropy reflect epileptic discharges: The first discharge was spontaneous (•) and the three following (•) were provoked by hyperventilation (•). The second interval of low entropy precedes a period of slowly decreasing entropy. This is related to a deceleration of the EEG, being a physiological consequence of the hyperventilation.

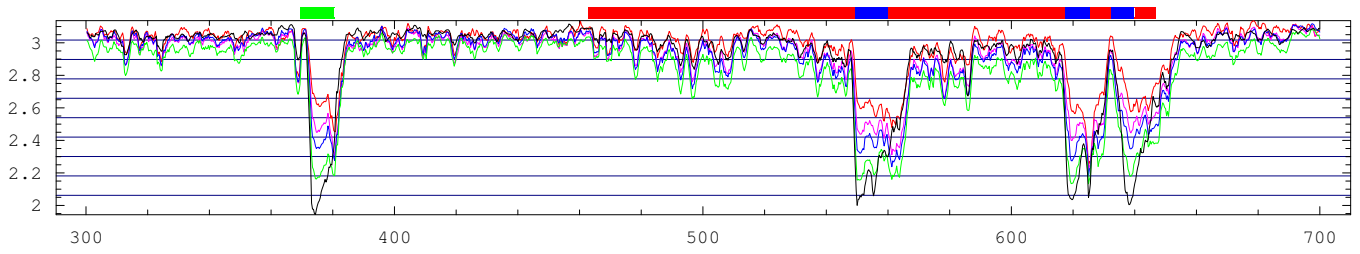

Fig. 2. Pooled permutation entropies for different delays.

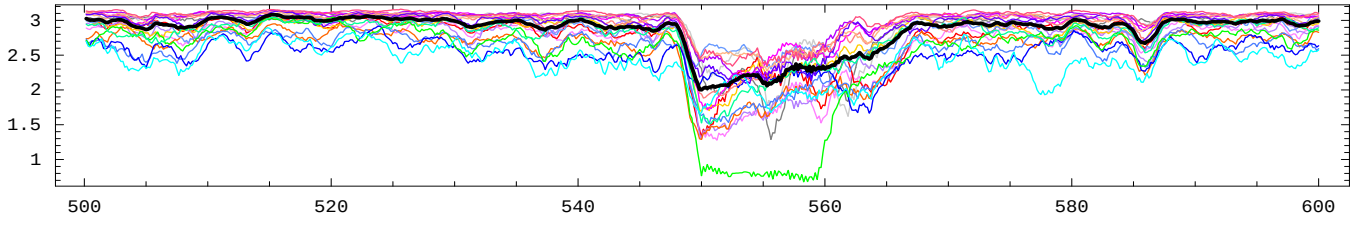

Fig. 3. Pooled and channel permutation entropies.

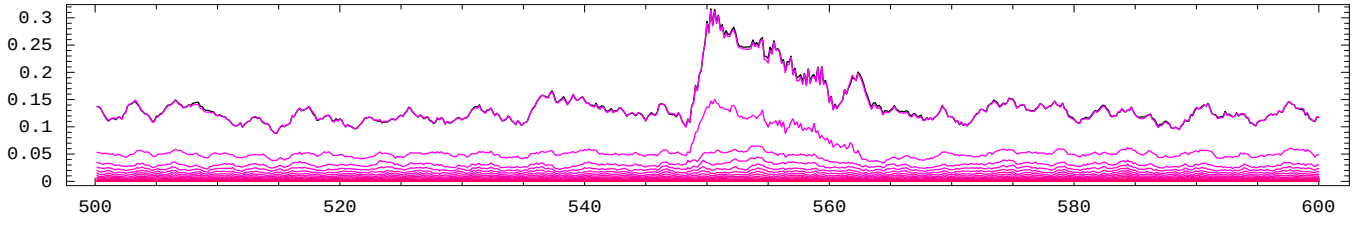

Fig. 4. Parts of contingency and mutual information.

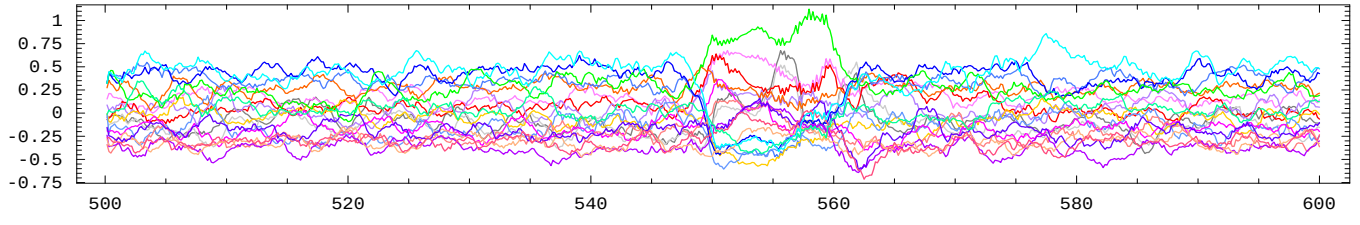

Fig. 5. Time-dependent "vertical" CA-Plot.

Figures 3–5 provide more details concerning the second epileptic period and some time before and after it for delay 1, the most sensitive delay among the considered ones. The channel permutation entropies are drawn in contrast to the pooled permutation entropy (thick black curve) in Fig. 3 (compare Fig. 1), and Fig. 4 shows the contingency ϕ_t^2 (first curve from above) and its parts $\rho_t^{(1)}$ not being explained by dimension 1 (second curve from above). The explained portion of the whole contingency lies between $\frac{1}{2}$ and $\frac{2}{3}$ at nearly all time points. Additionally, the parts of contingency not explained by dimensions $2, 3, \ldots$ are indicated in Fig. 4.

We have included a black curve into Fig. 4 being partially covered by the contingency curve

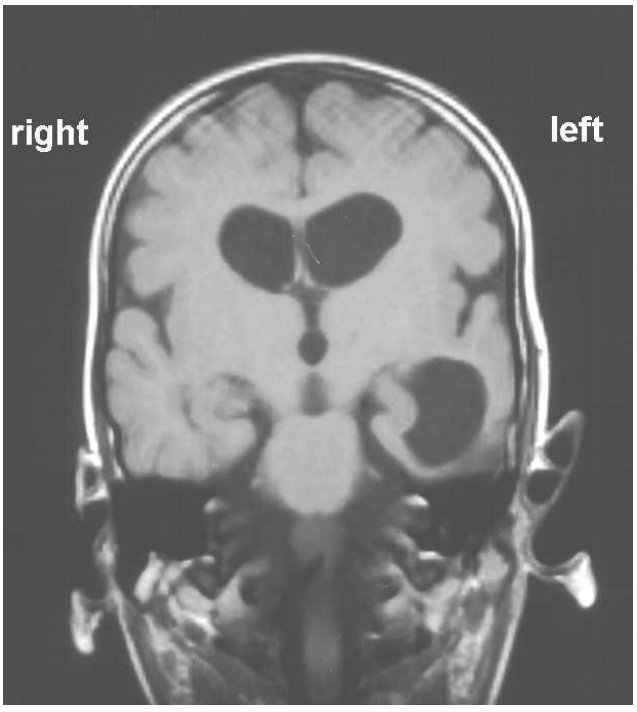

Fig. 6. MRI.

and showing the double mutual information between channels and ordinal patterns. We will discuss the relation of contingency and mutual information in the following section. Here we note that the almost coincidence of the two curves, which we have obtained in all numerical experiments, allows to consider formula (1) in the case of EEG-data.

The one-dimensional CA-Plot reflecting "distances" between the channels is given in Fig. 5. For each time the one-dimensional approximation of the ensemble of channel profiles is drawn in vertical direction. One gets an impression which channels are "near" respectively "far" from each other.

Example 2. Figures 8–12 show numerical results obtained for a boy at an age between 8 to 11 years with connatal toxoplasmosis for $\tau = 4$. Here the MRI has revealed lesions predominatly in the left temporal lobe, as seen in Fig. 6. The permutation entropy plots given in Fig. 8 are calculated from an EEG obtained at the age of 8 years. One can see that the entropy curves are strongly fluctuating,

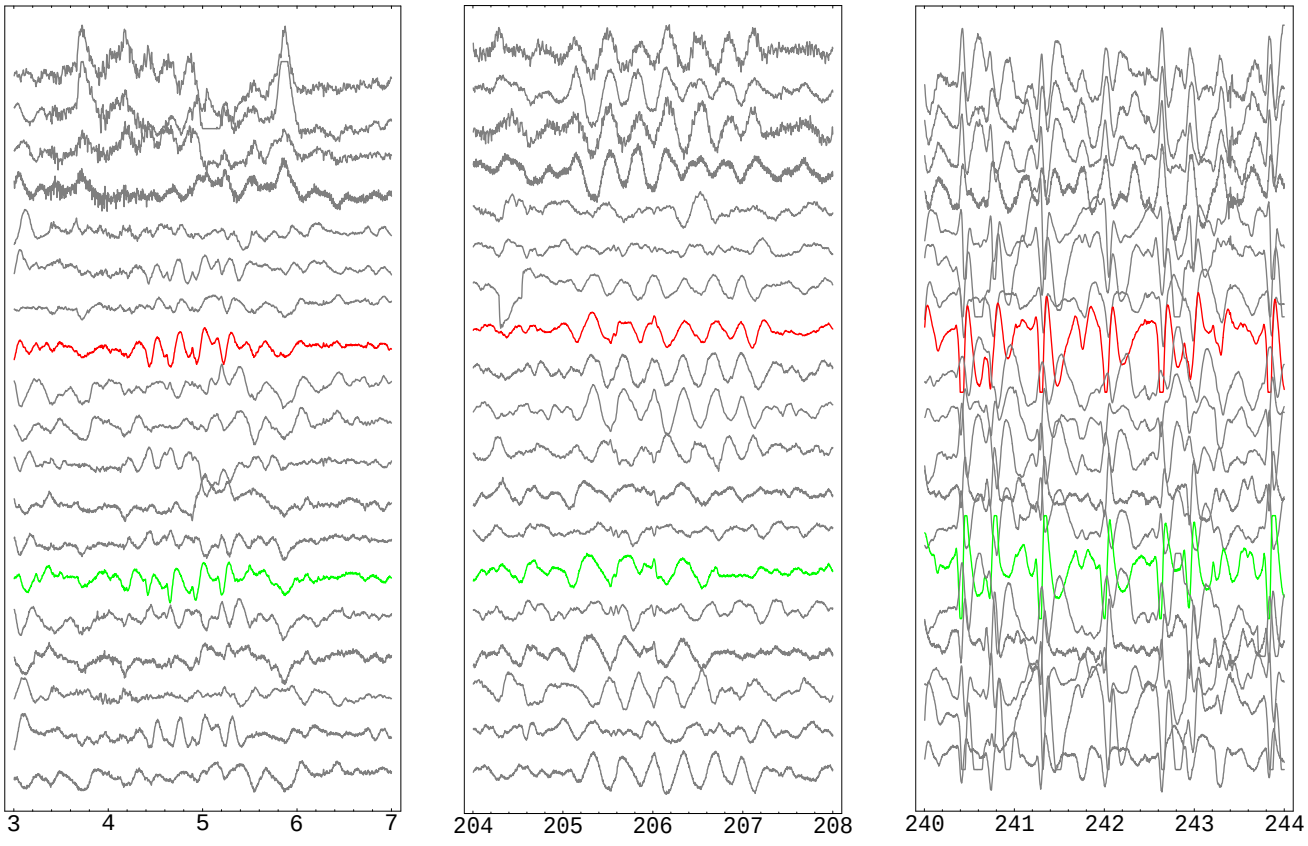

Fig. 7. Original EEG-data.

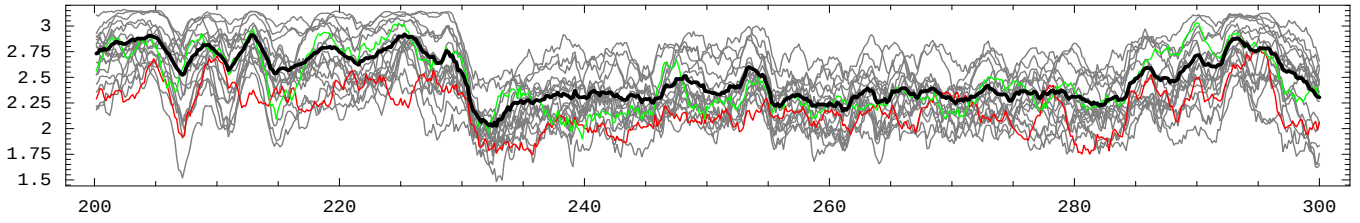

Fig. 8. At an age of 8 years.

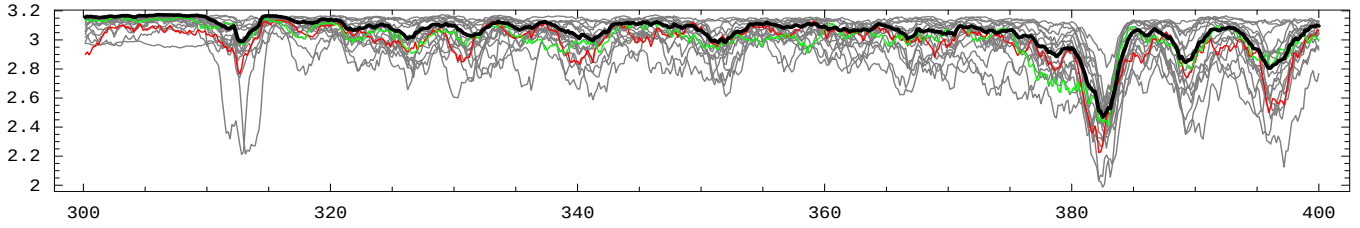

Fig. 9. Before implantation.

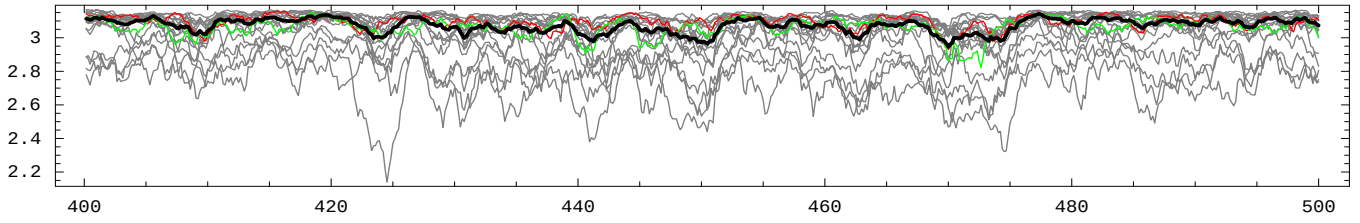

Fig. 10. One month after implantation.

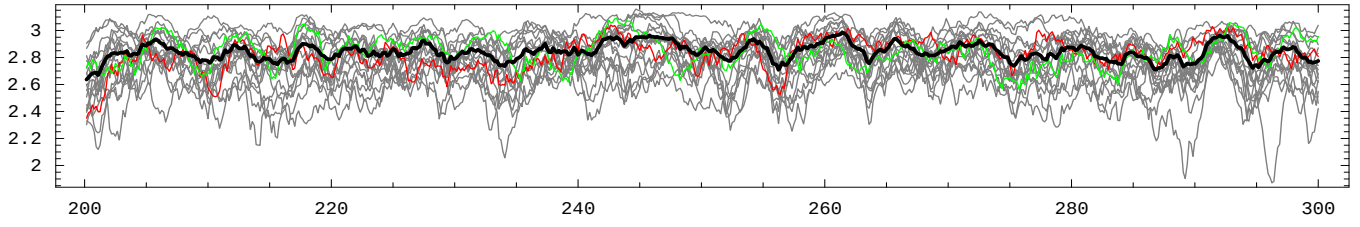

Fig. 11. Four months after implantation.

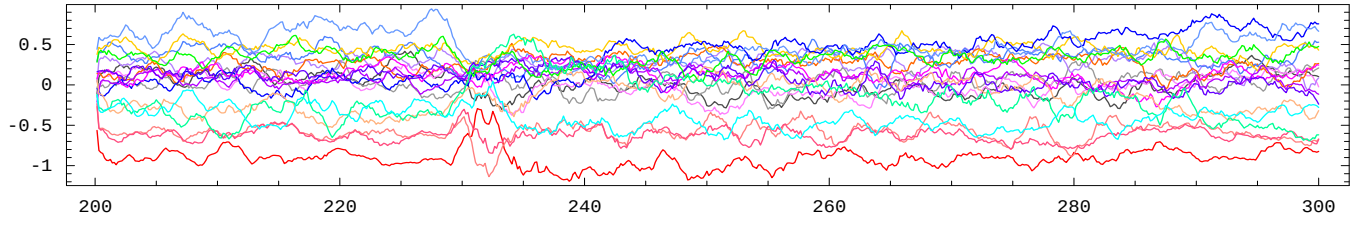

Fig. 12. "Separated" channels.

and that T3 and P3 provide low permutation entropies relative to the whole ensemble of entropies, where P3 is substantially below the pooled entropies (compare formula (1)).

This is in concordance with the fact that the EEG contains much epileptic activity, with sources in P3 and T3. For example, there is an epileptic seizure beginning at 228 s. We have added three short segments of the EEG, represented by Fig. 7. The segment on the left shows "sharp wave series" in P3 and T3, indicating epileptic sources, and behind the right part of the graphics is a generalized seizure. Note that the entropies at P3 and partially at T3 are low both in epileptic and in nonepileptic periods, which need not be immediately visible in the original EEG (compare the central and right parts of Fig. 7).

At the age of 11 years the boy was implanted a vagus nerve stimulator, with the aim to reduce the frequency and intensity of seizures. Figure 9 shows the permuation entropies for an EEG derived immediately before the implantation of the stimulator, and Figs. 10 and 11 display the results from EEG's derived one and four months after the implantation. Whereas the results illustrated in Fig. 9 are similar to those given in Fig. 8, the channel permutation entropies of P3 and T3 provided in Figs. 10 and 11 are nearly on the same level as the pooled permutation entropies.

Figure 12 drawn from the data at the age of 8 years was included in order to show that the time-dependent CA-plots can look very different. As for Fig. 5, we have chosen the delay $\delta = 1$, the arrangement of "channel projections" in Fig. 12 however is much more fixed in time. This could indicate more local separation of brain activity. Also note that the portion of contingency "explained in Fig. 12" is a bit less than 3/4 for the most time points.

5. Some Theoretical Background

5.1. *Entropy, mutual information and contingency*

In order to have a closer look at formula (1), let $i = 1, 2, \ldots, m$ and $j = 1, 2, \ldots, n$ be the categories of two variables ξ and η, respectively, and consider the matrix $P = (p_{ij})_{i,j=1}^{m,n}$, where p_{ij} denote the frequencies of combinations (i, j) relative to N given observations. Further, let $p_{i\cdot} = \sum_{j=1}^{m} p_{ij}$ and $p_{\cdot j} = \sum_{i=1}^{n} p_{ij}$ be the marginal relative frequencies related to P and assume that they are different from 0 for all categories i, j.

For $P = P_t$ as given in Sec. 3, ξ acts as a "channel" variable and η as an "ordinal pattern" variable. In this case $N = \delta m$ and $p_{i\cdot} = \frac{1}{m}$ for all i. We, however, want to consider the general situation. This does not need some additional effort; on the contrary, the representation is a bit easier when letting the symmetry of the role of the $p_{i\cdot}$ and the $p_{\cdot j}$ visible. (For a more detailed discussion of the following concepts, we refer to [Rényi, 1977].)

Remark. We want to emphasize, that in the present note we adopt the viewpoint of only exploring empirical symbolic data obtained from a system (but there is a "modeling" approach that is thinkable). So the subsequent representation starts from an empirical (!) distribution of ξ and η.

Consider the three (empirical) Shannon *entropies* $H(\xi) = -\sum_{i=1}^{m} p_{i\cdot} \ln p_{i\cdot}$, $H(\eta) = -\sum_{j=1}^{n} p_{\cdot j} \ln p_{\cdot j}$ and $H(\xi, \eta) = -\sum_{i,j=1}^{m,n} p_{ij} \ln p_{ij}$ measuring the "mean amounts of incertainty" for the "outcomes" of the variables ξ, η, and the paired variable (ξ, η). Note that the entropy of a nominal variable ξ which provides m different categories is maximal with value $H(\xi) = \ln m$ if the categories are equidistributed (as for $P = P_t$).

The degree of independency between ξ and η can be measured by the *contingency* $\varphi^2(\xi, \eta) = \sum_{i,j=1}^{m,n} \frac{(p_{ij} - p_{i\cdot} p_{\cdot j})^2}{p_{i\cdot} p_{\cdot j}}$ and the *mutual information* $MI(\xi, \eta) = H(\xi) + H(\eta) - H(\xi, \eta) = \sum_{i,j=1}^{m,n} p_{ij} \log \left(\frac{p_{ij}}{p_{i\cdot} p_{\cdot j}} \right)$. Their theoretical analogues (for random variables ξ, η and probabilities p_{ij}) are vanishing if ξ and η are statistically independent. Moreover, it is well known that $2NMI$ and $N\varphi^2$ — considered as "sample functions" depending on N random combinations (i, j) — have approximately the same χ^2-distributions for large N under appropriate model (!) assumptions and under the hypothesis of independency (see e.g. [Andersen, 1994]).)

For all EEG-data considered we found that the contingency and the double mutual information were nearly the same at all times. For an illustration of this fact, which was mentioned in Sec. 4, see Fig. 4. The black curve showing the double mutual information is nearly covered by the contingency curve. Generally, if $\varphi^2(\xi, \eta) \approx 2MI(\xi, \eta)$, then

$$H(\eta) = H(\xi, \eta) - H(\xi) + MI(\xi, \eta)$$
$$\approx H(\xi, \eta) - H(\xi) + \frac{\varphi^2(\xi, \eta)}{2}$$

$$= \sum_{i=1}^{m} p_{i\cdot} \left(\sum_{j=1}^{n} \frac{p_{ij}}{p_{i\cdot}} \ln \frac{p_{ij}}{p_{i\cdot}} \right) + \frac{\varphi^2(\xi,\eta)}{2} .$$

This provides formula (1) in the (EEG) case that $p_{i\cdot} = \frac{1}{m}$ for all $i = 1, 2, \ldots, m$.

5.2. *Relation of contingency and mutual information*

In order to give a relation between mutual information and contingency, let us show the following equality:

Proposition 1. *If $p_{i\cdot}, p_{\cdot j} \neq 0$ for all $i = 1, 2, \ldots, m$ and $j = 1, 2, \ldots, n$, then the following is valid:*

$$\begin{aligned} 2MI(\xi,\eta) = \varphi^2(\xi,\eta) &+ \sum_{\substack{i,j: \\ p_{ij}=0}} p_{i\cdot} p_{\cdot j} \\ &+ \sum_{\substack{i,j: \\ 0 \neq p_{ij} < p_{i\cdot} p_{\cdot j}}} p_{ij} \sum_{k=3}^{\infty} \frac{k-2}{k} \left(1 - \frac{p_{ij}}{p_{i\cdot} p_{\cdot j}} \right)^k \\ &- \sum_{\substack{i,j: \\ p_{ij} > p_{i\cdot} p_{\cdot j}}} p_{ij} \sum_{k=3}^{\infty} \frac{k-2}{k} \left(1 - \frac{p_{i\cdot} p_{\cdot j}}{p_{ij}} \right)^k . \end{aligned} \tag{4}$$

Proof. $MI(\xi,\eta) = H(\xi) + H(\eta) - H(\xi,\eta) = \sum_{i,j} q_{ij}$ holds with $q_{ij} = p_{ij} \ln \frac{p_{ij}}{p_{i\cdot} p_{\cdot j}}$. In order to show (4), we compare the terms $2q_{ij}$ and $\frac{(p_{ij} - p_{i\cdot} p_{\cdot j})^2}{p_{i\cdot} p_{\cdot j}}$. Both are vanishing if $p_{ij} = p_{i\cdot} p_{\cdot j}$, and in the case $p_{ij} = 0$ we have

$$2q_{ij} = 0 = \frac{(p_{ij} - p_{i\cdot} p_{\cdot j})^2}{p_{i\cdot} p_{\cdot j}} + 2p_{ij} - p_{i\cdot} p_{\cdot j} . \tag{5}$$

So the more interesting cases are $0 \neq p_{ij} < p_{i\cdot} p_{\cdot j}$ and $p_{ij} > p_{i\cdot} p_{\cdot j}$, where we define $r_{ij} := 1 - \frac{p_{ij}}{p_{i\cdot} p_{\cdot j}}$ and $s_{ij} := 1 - \frac{p_{i\cdot} p_{\cdot j}}{p_{ij}}$, respectively. By use of the Taylor expansion $\ln(1-x) = -\sum_{k=1}^{\infty} \frac{1}{k} x^k$ of the logarithm for $x = r_{ij}$ and $x = s_{ij}$, we show

$$\begin{aligned} 2q_{ij} &= \frac{(p_{ij} - p_{i\cdot} p_{\cdot j})^2}{p_{i\cdot} p_{\cdot j}} + 2p_{ij} - 2p_{i\cdot} p_{\cdot j} \\ &\quad + p_{ij} \sum_{k=3}^{\infty} \frac{k-2}{k} r_{ij}^k \end{aligned} \tag{6}$$

for $0 \neq p_{ij} < p_{i\cdot} p_{\cdot j}$ and

$$\begin{aligned} 2q_{ij} &= \frac{(p_{ij} - p_{i\cdot} p_{\cdot j})^2}{p_{i\cdot} p_{\cdot j}} + 2p_{ij} - 2p_{i\cdot} p_{\cdot j} \\ &\quad - p_{ij} \sum_{k=3}^{\infty} \frac{k-2}{k} s_{ij}^k \end{aligned} \tag{7}$$

for $p_{ij} > p_{i\cdot} p_{\cdot j}$. Then (4) follows by summing up Eqs. (5)–(7).

In the case $0 \neq p_{ij} < p_{i\cdot} p_{\cdot j}$ we obtain

$$\begin{aligned} 2q_{ij} &= -\sum_{k=1}^{\infty} \frac{2}{k} p_{ij} r_{ij}^k \\ &= 2p_{ij} \frac{p_{ij} - p_{i\cdot} p_{\cdot j}}{p_{i\cdot} p_{\cdot j}} - \sum_{k=2}^{\infty} \frac{2}{k} p_{ij} r_{ij}^k \\ &= 2p_{i\cdot} p_{\cdot j} \frac{p_{ij}^2 - p_{ij}\, p_{i\cdot} p_{\cdot j}}{p_{i\cdot}^2 p_{\cdot j}^2} - \sum_{k=2}^{\infty} \frac{2}{k} p_{ij} r_{ij}^k \\ &= 2p_{i\cdot} p_{\cdot j} \left(\frac{p_{ij} - p_{i\cdot} p_{\cdot j}}{p_{i\cdot} p_{\cdot j}} \right)^2 - 2p_{i\cdot} p_{\cdot j} \left(\frac{p_{i\cdot}^2 p_{\cdot j}^2 - p_{ij} p_{i\cdot} p_{\cdot j}}{p_{i\cdot}^2 p_{\cdot j}^2} \right) - \sum_{k=2}^{\infty} \frac{2}{k} p_{ij} r_{ij}^k \\ &= \frac{(p_{ij} - p_{i\cdot} p_{\cdot j})^2}{p_{i\cdot} p_{\cdot j}} + 2p_{ij} - 2p_{i\cdot} p_{\cdot j} + p_{i\cdot} p_{\cdot j} r_{ij}^2 - \sum_{k=2}^{\infty} p_{ij} r_{ij}^k + \sum_{k=3}^{\infty} \frac{k-2}{k} p_{ij} r_{ij}^k . \end{aligned} \tag{8}$$

Since $p_{i\cdot} p_{\cdot j} r_{ij}^l - p_{ij} r_{ij}^l = p_{i\cdot} p_{\cdot j} r_{ij}^{l+1}$ for all $l = 2, 3, 4, \ldots$, one easily shows $p_{i\cdot} p_{\cdot j} r_{ij}^2 - \sum_{k=2}^{\infty} p_{ij} r_{ij}^k = 0$ by using induction. Thus (8) turns into Eq. (6).

In the case $p_{ij} > p_{i\cdot} p_{\cdot j}$ we get

$$2q_{ij} = -2p_{ij} \ln \frac{p_{i\cdot} p_{\cdot j}}{p_{ij}} = \sum_{k=1}^{\infty} \frac{2}{k} p_{ij} s_{ij}^k = 2p_{ij} - 2p_{i\cdot} p_{\cdot j} + \frac{(p_{ij} - p_{i\cdot} p_{\cdot j})^2}{p_{ij}} + \sum_{k=3}^{\infty} \frac{2}{k} p_{ij} s_{ij}^k .$$

From this and the equality

$$\frac{(p_{ij}-p_{i\cdot}p_{\cdot j})^2}{p_{ij}}-\frac{(p_{ij}-p_{i\cdot}p_{\cdot j})^2}{p_{i\cdot}p_{\cdot j}}=-\frac{(p_{ij}-p_{i\cdot}p_{\cdot j})^3}{p_{ij}p_{i\cdot}p_{\cdot j}}=-\frac{p_{ij}^2}{p_{i\cdot}p_{\cdot j}}\,s_{ij}^3$$

it follows

$$\begin{aligned}2q_{ij}&=\frac{(p_{ij}-p_{i\cdot}p_{\cdot j})^2}{p_{i\cdot}p_{\cdot j}}+2p_{ij}-2p_{i\cdot}p_{\cdot j}-\frac{p_{ij}^2}{p_{i\cdot}p_{\cdot j}}\,s_{ij}^3+\sum_{k=3}^{\infty}\frac{2}{k}\,p_{ij}s_{ij}^k\\&=\frac{(p_{ij}-p_{i\cdot}p_{\cdot j})^2}{p_{i\cdot}p_{\cdot j}}+2p_{ij}-2p_{i\cdot}p_{\cdot j}-\frac{p_{ij}^2}{p_{i\cdot}p_{\cdot j}}\,s_{ij}^3+\sum_{k=3}^{\infty}p_{ij}s_{ij}^k-\sum_{k=3}^{\infty}\frac{k-2}{k}\,p_{ij}s_{ij}^k.\end{aligned}$$

Now $p_{ij}s_{ij}^l-\frac{p_{ij}^2}{p_{i\cdot}p_{\cdot j}}s_{ij}^l=-\frac{p_{ij}^2}{p_{i\cdot}p_{\cdot j}}s_{ij}^{l+1}$ for $l=3,4,5,\ldots$ implies

$$-\frac{p_{ij}^2}{p_{i\cdot}p_{\cdot j}}\,s_{ij}^3+\sum_{k=3}^{\infty}p_{ij}s_{ij}^k=0\,,$$

showing Eq. (7). ■

Equality (4) and simple geometric series calculations provide the unequalities

$$\begin{aligned}2MI(\xi,\eta)-\varphi^2(\xi,\eta)&\le\sum_{\substack{i,j:\\p_{ij}=0}}p_{i\cdot}p_{\cdot_j}+\sum_{\substack{i,j:\\0\neq p_{ij}<p_{i\cdot}p_{\cdot j}}}p_{ij}\left(1-\frac{p_{ij}}{p_{i\cdot}p_{\cdot j}}\right)^3\left(\frac{1}{3}+\left(1-\frac{p_{ij}}{p_{i\cdot}p_{\cdot j}}\right)\sum_{k=0}^{\infty}\left(1-\frac{p_{ij}}{p_{i\cdot}p_{\cdot j}}\right)^k\right)\\&=\sum_{\substack{i,j:\\p_{ij}=0}}p_{i\cdot}p_{\cdot_j}+\sum_{\substack{i,j:\\0\neq p_{ij}<p_{i\cdot}p_{\cdot j}}}p_{ij}\left(1-\frac{p_{ij}}{p_{i\cdot}p_{\cdot j}}\right)^3\left(\frac{p_{i\cdot}p_{\cdot j}}{p_{ij}}-\frac{2}{3}\right)\\&=\sum_{\substack{i,j:\\p_{ij}=0}}p_{i\cdot}p_{\cdot_j}+\sum_{\substack{i,j:\\0\neq p_{ij}<p_{i\cdot}p_{\cdot j}}}\frac{(p_{ij}-p_{i\cdot}p_{\cdot j})^2}{p_{i\cdot}p_{\cdot j}}\left(1-\frac{p_{ij}}{p_{i\cdot}p_{\cdot j}}\right)\left(1-\frac{2}{3}\,\frac{p_{ij}}{p_{i\cdot}p_{\cdot j}}\right)\end{aligned}\tag{9}$$

and

$$\begin{aligned}\varphi^2(\xi,\eta)-2MI(\xi,\eta)&\le\sum_{\substack{i,j:\\p_{ij}>p_{i\cdot}p_{\cdot j}}}p_{ij}\left(1-\frac{p_{i\cdot}p_{\cdot j}}{p_{ij}}\right)^3\left(\frac{1}{3}+\left(1-\frac{p_{i\cdot}p_{\cdot j}}{p_{ij}}\right)\sum_{k=0}^{\infty}\left(1-\frac{p_{i\cdot}p_{\cdot j}}{p_{ij}}\right)^k\right)\\&=\sum_{\substack{i,j:\\p_{ij}>p_{i\cdot}p_{\cdot j}}}p_{ij}\left(1-\frac{p_{i\cdot}p_{\cdot j}}{p_{ij}}\right)^3\left(\frac{p_{ij}}{p_{i\cdot}p_{\cdot j}}-\frac{2}{3}\right)\\&=\sum_{\substack{i,j:\\p_{ij}>p_{i\cdot}p_{\cdot j}}}\frac{(p_{ij}-p_{i\cdot}p_{\cdot j})^2}{p_{i\cdot}p_{\cdot j}}\left(1-\frac{p_{i\cdot}p_{\cdot j}}{p_{ij}}\right)\left(1-\frac{2}{3}\,\frac{p_{i\cdot}p_{\cdot j}}{p_{ij}}\right).\end{aligned}\tag{10}$$

If $1-\varepsilon\le\frac{p_{ij}}{p_{i\cdot}p_{\cdot j}}\le\frac{1}{1-\varepsilon}$ for some positive $\varepsilon<1$, then $1-\frac{p_{ij}}{p_{i\cdot}p_{\cdot j}},1-\frac{p_{i\cdot}p_{\cdot j}}{p_{ij}}\le\varepsilon,\frac{p_{i\cdot}p_{\cdot j}}{p_{ij}}-\frac{2}{3},\frac{p_{ij}}{p_{i\cdot}p_{\cdot j}}-\frac{2}{3}\le\frac{1+2\varepsilon}{3(1-\varepsilon)}$ and $1-\frac{2}{3}\frac{p_{ij}}{p_{i\cdot}p_{\cdot j}},1-\frac{2}{3}\frac{p_{i\cdot}p_{\cdot j}}{p_{ij}}\le\frac{1+2\varepsilon}{3}$. From this, (9) and (10), and $\sum_{i,j}p_{ij}=1$ and $\sum_{i,j}\frac{(p_{ij}-p_{i\cdot}p_{\cdot j})^2}{p_{i\cdot}p_{\cdot j}}=\varphi^2(\xi,\eta)$ one easily obtains the following relations between $MI(\xi,\eta)$ and $\varphi^2(\xi,\eta)$:

Corollary 2. *If $1-\varepsilon\le\frac{p_{ij}}{p_{i\cdot}p_{\cdot j}}\le\frac{1}{1-\varepsilon}$ for all i,j and for some ε with $0<\varepsilon<1$, then the following is valid:*

(i) $|2MI(\xi,\eta)-\varphi^2(\xi,\eta)|\le\varepsilon^3(\frac{1}{3}+\frac{\varepsilon}{1-\varepsilon})$.
(ii) $|\frac{2MI(\xi,\eta)}{\varphi^2(\xi,\eta)}-1|\le\varepsilon\frac{1+2\varepsilon}{3}$.

Corollary 2 suggests why mutual information and contingency behave rather similarly if ξ and η are not "too dependent". Whereas (i) provides bounds for the difference of double mutual information and contingency, (ii) estimates their distance in the

multiplicative sense. However, the inequalities obtained are usually very rough, mainly since parts of the "positive" and "negative" sums in (4) cancel each other out.

For $p_i = \frac{1}{m}$, as we have in the EEG case, $\frac{p_{ij}}{p_{i\cdot}p_{\cdot j}} = \frac{mp_{ij}}{p_{\cdot j}}$ holds. Here the assumption of Corollary 2 is equivalent to $(1-\varepsilon)p_{\cdot j} \le mp_{ij} \le \frac{p_{\cdot j}}{(1-\varepsilon)}$, roughly saying that all "channel" distributions $\mathbf{r}_i$ are near to the "pooled" distribution $\mathbf{r}$ or, equivalently, are not too different from each other.

References

Accardo, A., Affinito, M., Carrozzi, M. & Bouquet, F. [1997] "Use of the fractal dimension for the analysis of electroencephalographic time series," *Biol. Cybern.* **77**, 335–350.

Andersen, E. B. [1994] *The Statistical Analysis of Categorical Data* (Springer-Verlag, Berlin, Heidelberg, NY).

Bandt, C. & Pompe, B. [2002] "Permutation entropy — a natural complexity measure for time series," *Phys. Rev. Lett.* **88**, 174102.

Bandt, C. Keller, G. & Pompe, B. [2002] "Entropy of interval maps via permutations," *Nonlinearity* **15**, 1595–1602.

Benzécri, J.-P. [1973] *L'analyse des Données* (Dunod, Paris).

Benzécri, J.-P. [1992] *Correspondence Analysis Handbook* (Marcel Dekker, NY).

Blume, W. T. & Kaibara, M. [1999] *Atlas of Pediatric Electroencephalography* (Lippincott-Raven, Philadelphia).

Daw, C. S., Finney, C. E. A., Nguyen, K. & Halow, J. S. [1998] "Symbol statistics: A new tool for understanding multiphase flow phenomena," *Int. Mechanical Engineering Congress & Exposition*, Anaheim, CA, pp. 405–411.

Ebeling, W. & Nicolis, G. [1992] "Word frequency and entropy of symbolic sequences: A dynamical perspective," *Chaos Solit. Fract.* **6**, 635–650.

Finney, C. E. A., Green, J. B. Jr. & Daw, C. S. [1998] "Symbolic time-series analysis of engine combustion measurements," *SAE* Paper No. 980624.

Grassberger, P. & Procaccia, I. [1983] "Characterization of strange attractors," *Phys. Rev. Lett.* **50**, 346–349.

Greenacre, M. J. [1984] *Theory and Applications of Correspondence Analysis* (Academic Press, London).

Hao, B.-L. & Zheng, W. M. [1998] *Applied Symbolic Dynamics and Chaos* (World Scientific, Singapore).

Herzel, H.-P. & Große, I. [1995] "Measuring correlations in symbol sequences," *Physica* **A4**, 518–542.

Lehnertz, K., Arnhold, J. Grassberger, P. & Elger, C. E. (eds.) [2000] *Chaos in Brain?* (World Scientific, Singapore).

Rényi, A. [1977] *Wahrscheinlichkeitsrechnung: Mit einem Anhang über Informationstheorie* (Deutscher Verlag der Wissenschaften, Berlin).

Schwarz, U., Benz, A. O., Kurths, J. & Witt, A. [1993] "Analysis of solar spike events by means of symbolic dynamics methods," *Astro. Astrophys.* **277**, 215–224.

Takens, F. [1981] "Detecting strange attractors in turbulence," in *Dynamical Systems and Turbulence*, Warwick 1980, Coventry, 1979/1980 (Springer-Verlag, Berlin).

Theiler, J. [1995] "On the evidence for low-dimensional chaos in an epileptic electroencephalogram," *Phys. Lett.* **A196**, 335–341.

Wittfeld, K. [2001] *Die Geometrie der Korrespondenzanalyse*, Diploma thesis (Greifswald).

Appendix

In order to let the reader know how the one-dimensional CA-plots used above can be determined in a very fast manner, we give some more details concerning Correspondence Analysis (CA). The starting point is a matrix P of relative frequencies p_{ij} as given at the beginning Sec. 5. More general than in the time series case, we consider the distributions $\mathbf{r}_i = (\frac{p_{ij}}{p_{i\cdot}})_{j=1}^n; i = 1, 2, \ldots, m$. This is useful since we did not find references for the special case that $p_{i\cdot} = \frac{1}{m}$ for all i. For a comprehensive discussion of the following we refer to [Greenacre, 1984].

In texts on CA the space $\mathbb{R}^n$ is usually equipped with a norm $\|\cdot\|_{D_c}$ different from $\|\cdot\|$ given in Sec. 3. This norm, respecting the weights given by $p_{\cdot j}; j = 1, 2, \ldots, n$ — components with higher frequency are more important — is deduced from the scalar product $\langle \mathbf{s}, \tilde{\mathbf{s}} \rangle_{D_c} = \mathbf{s} D_c \tilde{\mathbf{s}}$, where D_c denotes the $n \times n$ (column sum) diagonal matrix with diagonal $p_{\cdot 1}, p_{\cdot 2}, \ldots, p_{\cdot n}$. On $\mathbb{R}^m$ we need the norm $\|\cdot\|_{D_r}$, defined similarly as $\|\cdot\|_{D_c}$ for the (row sum) diagonal matrix D_r with diagonal entries $p_{1\cdot}$, $p_{2\cdot}, \ldots, p_{m\cdot}$.

One easily shows that the map I defined by

$$I((s_j)_{j=1}^n) = (s_j/p_{\cdot j} - 1)_{j=1}^n$$

forms an isometry from $(\mathbb{R}^n, \|\cdot\|)$ onto $(\mathbb{R}^n, \|\cdot\|_{D_c})$. So the problem of an "optimal" l-dimensional approximation of $\mathbf{r}_i$ in $(\mathbb{R}^n, \|\cdot\|)$ is equivalent to that of an "optimal" l-dimensional approximation of the *profiles* $I(\mathbf{r}_i) = (\frac{p_{ij}}{p_{i\cdot}p_{\cdot j}} - 1)_{j=1}^n$ in $(\mathbb{R}^n, \|\cdot\|_{D_c})$. It is plausible to consider an l-dimensional affine subspace S of $\mathbb{R}^n$ as *optimal* (for dimension l) if the sum of its squared distances to the $\mathbf{r}_i$ weighted with $p_{i\cdot}$ is minimal.

Such an optimal subspace S contains the "mean profile" $I(\mathbf{r}) = 0$, hence it is a linear one, and if $\mathbf{v}^S$ denotes the projection of a vector $\mathbf{v}$ onto S, the sum $\sum_{i=1}^{m} p_{i\cdot}\|I(\mathbf{r}_i) - I(\mathbf{r}_i)^S\|_{D_c}^2$ is minimal for dimension l. Since the contingency can be considered as a weighted sum of the squared lengths of all profiles $I(\mathbf{r}_i)$, one has the following orthogonal decomposition of ϕ^2 (compare Sec. 3):

$$\begin{aligned}\varphi^2(\xi,\eta) &= \sum_{i,j=1}^{m,n} \frac{(p_{ij} - p_{i\cdot}p_{\cdot j})^2}{p_{i\cdot}p_{\cdot j}} \\ &= \sum_{i=1}^{m} p_{i\cdot}\|I(\mathbf{r}_i)\|_{D_c}^2 \\ &= \sum_{i=1}^{m} p_{i\cdot}\|I(\mathbf{r}_i)^S\|_{D_c}^2 \\ &\quad + \sum_{i=1}^{m} p_{i\cdot}\|I(\mathbf{r}_i) - I(\mathbf{r}_i)^S\|_{D_c}^2 . \qquad \text{(A.1)}\end{aligned}$$

One approach to finding optimal subspaces is the application of Singular Value Decomposition (SVD) to the matrix $A = (\frac{p_{ij}}{p_{i\cdot}p_{\cdot j}} - 1)_{i,j=1}^{m,n}$. Its rows are the profiles $I(\mathbf{r}_i)$, hence the rank k of A is the dimension of the span of $\{\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_m\}$. SVD provides mutually orthogonal vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k \in (\mathbb{R}^n, \langle\cdot,\cdot\rangle_{D_c})$ of length 1, spanning the same linear subspace as $\{I(\mathbf{r}_1), I(\mathbf{r}_2), \ldots, I(\mathbf{r}_m)\}$, and strictly positive numbers $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_k$ not greater than 1, such that for $l = 1, 2, \ldots, k$ the following is valid:

(I) The span S of $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_l\}$ is optimal for dimension l with $\sum_{i=1}^{m} p_{i\cdot}\|I(\mathbf{r}_i)^S\|_{D_c}^2 = \lambda_1^2 + \lambda_2^2 + \ldots + \lambda_l^2$ (and $\sum_{i=1}^{m} p_{i\cdot}\|I(\mathbf{r}_i) - I(\mathbf{r}_i)^S\|_{D_c}^2 = \lambda_{1+1}^2 + \lambda_{l+2}^2 + \cdots + \lambda_k^2$),

(II) the "lth coordinate vector" $\mathbf{w}_l := (\langle I(\mathbf{r}_i), \mathbf{v}_l\rangle_{D_c})_{i=1}^{m} \in (\mathbb{R}^m, \langle\cdot,\cdot\rangle_{D_c})$ has length λ_l,

(III) $\mathbf{w}_l$ is an eigenvector corresponding to the eigenvalue λ_l^2 of the matrix $AD_cA^TD_r$.

Note that the directions defined by $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$ are unique if all λ_l are different, which is usually satisfied for empirical (noisy) data. Let us relate the above statements to the short discussion of CA in Sec. 3, where $P = P_t$, $p_{i\cdot} = \frac{1}{m}$ and $\varphi^2(\xi,\eta) = \varphi_t^2$. (For the general case, see [Greenacre, 1984].)

Since I is an isometry, the preimage $I^{-1}(S)$ of the span S of $\mathbf{v}_1$ is a one-dimensional affine subspace of $(\mathbb{R}^m, \|\cdot\|)$ supporting best possible one-dimensional approximations $\mathbf{r}_i^{(1)} = I^{-1}(I(\mathbf{r}_i)^S)$ of $\mathbf{r}_i$, and the decompositions of ϕ_t^2 according to (A.1) and (3) are equivalent. The latter and (I) imply $\rho^{(1)} = \lambda_2^2 + \lambda_3^2 + \cdots \lambda_k^2$. The vector $\mathbf{w}_1$ of dimension m consists of the first coordinates of the profiles $I(\mathbf{r}_i)$; $i = 1, 2, \ldots, m$ with respect to the system $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_l\}$, so it reflects them in $\mathbf{v}_1$-direction. Now we can say what the one-dimensional CA-Plot exactly is: It draws the components of $\mathbf{w}_1$ on a line.

(III) provides a simple method for obtaining $\mathbf{w}_1$. Since λ_1^2 is the leading eigenvalue of $O = AD_cA^TD_r = \frac{1}{m}AD_cA^T$, multiplication with O suppresses all directions orthogonal to $\mathbf{w}_1$. So by use of (II) and (III) one can easily see that an iterative application of the operator $\mathbf{w} \mapsto \frac{O(\mathbf{w})}{\|O(\mathbf{w})\|}\sqrt{\frac{\|O(\mathbf{w})\|}{\|\mathbf{w}\|}} = \frac{O(\mathbf{w})}{\sqrt{\|\mathbf{w}\|\,\|O(\mathbf{w})\|}}$ to a (random) vector approaches $\mathbf{w}_1$. In the time series context one can use that P_{t+k} is near P_t for small k, such that $\mathbf{w}_1$ at time t is a good start vector for approaching $\mathbf{w}_1$ at time $t + k$. This is the base for an extremely fast computing of the temporal one-dimensional CA-Plots given in Sec. 4.

Remark. (I)–(III) are also the base for an optimal l-dimensional representation of the system $\{\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_m\}$, which can easily be seen and which we do not want to discuss here. Only note that the part of contingency not explained by dimension l is $\lambda_{l+1}^2 + \lambda_{l+2}^2 + \cdots \lambda_k^2$ (according to (I)).